

Convergence Analysis of MCMC Algorithms for Bayesian Multivariate Linear Regression with Non-Gaussian Errors

James P. Hobert*, Yeun Ji Jung, Kshitij Khare and Qian Qin

Department of Statistics

University of Florida

Gainesville, FL 32611

May 2017

Abstract

Gaussian errors are sometimes inappropriate in a multivariate linear regression setting because, for example, the data contain outliers. In such situations, it is often assumed that the errors are iid from a distribution that is a scale mixture of multivariate normals. Combining this robust regression model with a default prior on the unknown parameters results in a highly intractable posterior density. Fortunately, there is a simple data augmentation (DA) algorithm and a corresponding Haar PX-DA algorithm that can be used to explore this posterior. This paper provides conditions (on the mixing density, h) for geometric ergodicity of the Markov chains underlying these Markov chain Monte Carlo (MCMC) algorithms. The main result is that, if h converges to 0 at the origin at an appropriate rate, and $\int_0^\infty u^{\frac{d}{2}} h(u) du < \infty$, where d is the dimension of the response, then both Markov chains are geometrically ergodic. This result is quite far-reaching. For example, it implies the geometric ergodicity of the DA and Haar PX-DA Markov chains whenever h is generalized inverse Gaussian, log-normal, inverted gamma (with shape parameter larger than $d/2$), or Fréchet (with shape parameter larger than $d/2$). The result also applies to certain subsets of the gamma, F , and Weibull families.

1 Introduction

Let Y_1, Y_2, \dots, Y_n be independent d -dimensional random vectors from the multivariate linear regression model

$$Y_i = \beta^T x_i + \Sigma^{\frac{1}{2}} \varepsilon_i, \quad (1)$$

Abbreviated title. Convergence of MCMC algorithms

Key words and phrases. Data augmentation algorithm, Drift condition, Geometric ergodicity, Haar PX-DA algorithm, Heavy-tailed distribution, Minorization condition, Scale mixture

where x_i is a $p \times 1$ vector of known covariates associated with Y_i , β is a $p \times d$ matrix of unknown regression coefficients, Σ is an unknown $d \times d$ positive definite scale matrix, and $\varepsilon_1, \dots, \varepsilon_n$ are iid d -dimensional errors. In situations where Gaussian errors are inappropriate, e.g., when the data contain outliers, scale mixtures of multivariate normal densities constitute a rich class of alternative error densities (see, e.g., Andrews and Mallows, 1974; Fernández and Steel, 1999, 2000; West, 1984). These mixtures take the form

$$f_h(\varepsilon) = \int_0^\infty \frac{u^{\frac{d}{2}}}{(2\pi)^{\frac{d}{2}}} \exp\left\{-\frac{u}{2}\varepsilon^T\varepsilon\right\} h(u) du,$$

where h is the density function of some positive random variable. We shall refer to h as a *mixing density*. By varying the mixing density, one can construct error densities with many different types of tail behavior. A well-known example is that when h is the density of a Gamma($\frac{\nu}{2}, \frac{\nu}{2}$) random variable, then f_h becomes the multivariate Student's t density with ν degrees of freedom, which, aside from a normalizing constant, is given by $[1 + \nu^{-1}\varepsilon^T\varepsilon]^{-\frac{d+\nu}{2}}$.

Let Y denote the $n \times d$ matrix whose i th row is Y_i^T , and let X stand for the $n \times p$ matrix whose i th row is x_i^T , and, finally, let ε represent the $n \times d$ matrix whose i th row is ε_i^T . Using this notation, we can state the n equations in (1) more succinctly as follows

$$Y = X\beta + \varepsilon\Sigma^{\frac{1}{2}}. \tag{2}$$

Let y and y_i denote the observed values of Y and Y_i , respectively.

Recent applications of the robust multivariate regression model (2) are mostly Bayesian in nature. These include Lee et al. (2016), who use a Bayesian version of (2) as a high-dimensional vector autoregressive (VAR) model, and Ala-Luhtala and Piché (2016), who develop a variational Bayes method for approximating the posterior distribution when there is ignorable missing data. There are also many recent papers that employ Bayesian models based on the univariate ($d = 1$) version of (2). Indeed, Nevo and Ritov (2016) use such a model in the high dimensional context where the number of predictors and the number of observations grow at a similar rate. These authors establish consistency of the Bayes estimator, and show that the posterior can be explored via Gibbs sampling. Another example is Yang and Yuan (2017), who develop a method of drawing iid samples from an approximation to the posterior distribution. Salazar et al. (2012) and Ferreira and Salazar (2014) develop default prior distributions for the specific case in which the errors are from the exponential power distribution (which corresponds to scaled positive stable mixing density), and Choy et al. (2016) perform a Bayesian analysis of insurance data.

In this paper, we consider a Bayesian version of (2) with an improper prior on (β, Σ) that takes the form $\omega(\beta, \Sigma) \propto |\Sigma|^{-a} I_{\mathcal{S}_d}(\Sigma)$ where $\mathcal{S}_d \subset \mathbb{R}^{\frac{d(d+1)}{2}}$ denotes the space of $d \times d$ positive definite matrices. Taking $a = (d + 1)/2$ yields the independence Jeffreys prior, which is a standard default prior for multivariate location scale problems, and is often used in conjunction with (2) (see, e.g.,

Fernández and Steel, 1999; Liu, 1996; Meng and van Dyk, 1999). The joint density of the data from model (2) is, of course, given by

$$f(y|\beta, \Sigma) = \prod_{i=1}^n \left[\int_0^\infty \frac{u^{\frac{d}{2}}}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} \exp \left\{ -\frac{u}{2} (y_i - \beta^T x_i)^T \Sigma^{-1} (y_i - \beta^T x_i) \right\} h(u) du \right]. \quad (3)$$

Define

$$m(y) = \int_{\mathcal{S}_d} \int_{\mathbb{R}^{p \times d}} f(y|\beta, \Sigma) \omega(\beta, \Sigma) d\beta d\Sigma.$$

The posterior distribution is proper precisely when $m(y) < \infty$. Let Λ denote the $n \times (p + d)$ matrix $(X : y)$. That is, Λ is the matrix that results when the $n \times d$ matrix y is appended to the right of the $n \times p$ matrix X . As we shall see, the following conditions are necessary for propriety:

(N1) $\text{rank}(\Lambda) = p + d$;

(N2) $n > p + 2d - 2a$.

We assume throughout the paper that (N1) and (N2) hold. Under these two conditions, the Markov chain of interest is well-defined, and we can engage in a convergence rate analysis whether the posterior is proper or not. This is a subtle point upon which we will expand in Section 3.

Of course, when the posterior is proper, it is given by

$$\pi^*(\beta, \Sigma|y) = \frac{f(y|\beta, \Sigma) \omega(\beta, \Sigma)}{m(y)}.$$

This density is (nearly always) intractable in the sense that posterior expectations cannot be computed in closed form. However, there is a well-known data augmentation algorithm (or two-variable Gibbs sampler) that can be used to explore this intractable posterior density (see, e.g., Liu, 1996). In order to state this algorithm, we must introduce some additional notation. For $z = (z_1, \dots, z_n)$, let Q be an $n \times n$ diagonal matrix whose i th diagonal element is z_i^{-1} . Also, define $\Omega = (X^T Q^{-1} X)^{-1}$ and $\mu = (X^T Q^{-1} X)^{-1} X^T Q^{-1} y$. We shall assume throughout the paper that

$$\int_0^\infty u^{\frac{d}{2}} h(u) du < \infty,$$

where h is the mixing density, and we will refer to this condition as “condition \mathcal{M} .” Finally, define a parametric family of univariate density functions indexed by $s \geq 0$ as follows

$$\psi(u; s) = b(s) u^{\frac{d}{2}} e^{-\frac{su}{2}} h(u),$$

where $b(s)$ is the normalizing constant. The data augmentation (DA) algorithm calls for draws from the inverse Wishart (IW_d) and matrix normal ($\text{N}_{p,d}$) distributions. The precise forms of the densities are given in the Appendix. We now present the DA algorithm. If the current state of the DA Markov chain is $(\beta_m, \Sigma_m) = (\beta, \Sigma)$, then we simulate the new state, $(\beta_{m+1}, \Sigma_{m+1})$, using the following three-step procedure.

Iteration $m + 1$ of the DA algorithm:

1. Draw $\{Z_i\}_{i=1}^n$ independently with $Z_i \sim \psi\left(\cdot; (\beta^T x_i - y_i)^T \Sigma^{-1} (\beta^T x_i - y_i)\right)$, and call the result $z = (z_1, \dots, z_n)$.

2. Draw

$$\Sigma_{m+1} \sim \text{IW}_d\left(n - p + 2a - d - 1, \left(y^T Q^{-1} y - \mu^T \Omega^{-1} \mu\right)^{-1}\right).$$

3. Draw $\beta_{m+1} \sim \text{N}_{p,d}(\mu, \Omega, \Sigma_{m+1})$

This algorithm is based on a so-called augmented posterior density, $\pi(\beta, \Sigma, z|y)$, where $z \in \mathbb{R}_+^n$ is a vector of latent data (formally defined in Section 2) such that $\int_{\mathbb{R}_+^n} \pi(\beta, \Sigma, z|y) dz = \pi^*(\beta, \Sigma|y)$. A single iteration consists of a draw from $\pi(z|\beta, \Sigma, y)$, followed by a draw from $\pi(\beta, \Sigma|z, y)$. The algorithm stated above has three steps (instead of two) because the draw from $\pi(\beta, \Sigma|z, y)$ is done sequentially, using $\pi(\Sigma|z, y)$ and $\pi(\beta|\Sigma, z, y)$.

Obviously, in order to run this DA algorithm, one must be able to make draws from $\psi(\cdot; s)$. When h is a standard density, ψ often turns out to be one as well. For example, when h is a gamma density, ψ is also gamma, and when h is inverted gamma, ψ is generalized inverse Gaussian (see Section 5). Even when ψ is not a standard density, it is still a simple entity - a univariate density on $(0, \infty)$ - and so is usually amenable to straightforward sampling. In particular, if it is possible to make draws from h , then h can be used as the candidate in a simple rejection sampler for ψ .

Denote the DA Markov chain by $\Phi = \{(\beta_m, \Sigma_m)\}_{m=0}^\infty$. The main contribution of this paper is to demonstrate that Φ is geometrically ergodic as long as h converges to zero at the origin at an appropriate rate. (A formal definition of geometric ergodicity is given in Section 3.) Our result is remarkable both for its simplicity and for its scope. Indeed, the conditions turn out to be extremely simple to check, and, at the same time, the result applies to a huge class of Monte Carlo Markov chains. It is well known among Markov chain Monte Carlo (MCMC) experts that establishing geometric ergodicity of practically relevant chains is extremely challenging. Thus, it is noteworthy that we are able to handle so many such chains simultaneously. Of course, the important practical and theoretical benefits of basing one's MCMC algorithm on a geometrically ergodic Markov chain have been well-documented by, e.g., Roberts and Rosenthal (1998), Jones and Hobert (2001), Flegal et al. (2008) and Łatuszyński et al. (2013). In order to give a precise statement of our main result, we now define three classes of mixing densities based on behavior near the origin.

Define $\mathbb{R}_+ = (0, \infty)$, and let $h : \mathbb{R}_+ \rightarrow [0, \infty)$ be a mixing density. If there is a $\delta > 0$ such that $h(u) = 0$ for all $u \in (0, \delta)$, then we say that h is *zero near the origin*. Now assume that h is strictly positive in a neighborhood of 0. If there exists a $c > -1$ such that

$$\lim_{u \rightarrow 0} \frac{h(u)}{u^c} \in \mathbb{R}_+,$$

then we say that h is *polynomial near the origin* with power c . Finally, if for every $c > 0$, there exists an $\eta_c > 0$ such that the ratio $\frac{h(u)}{u^c}$ is strictly increasing in $(0, \eta_c)$, then we say that h is *faster than polynomial near the origin*.

Every mixing density that is a member of a standard parametric family is either polynomial near the origin, or faster than polynomial near the origin. Indeed, the gamma, beta, F , Weibull, and shifted Pareto densities are all polynomial near the origin, whereas the inverted gamma, log-normal, generalized inverse Gaussian, and Fréchet densities are all faster than polynomial near the origin. We establish these facts in Section 5. Here is our main result.

Theorem 1. *Assume that (N1) and (N2) hold. Let h be a mixing density that satisfies condition \mathcal{M} . If any one of the following conditions holds, then the posterior distribution is proper and the DA Markov chain is geometrically ergodic.*

1. h is zero near the origin.
2. h is strictly positive in a neighborhood of zero, and faster than polynomial near the origin.
3. h is strictly positive in a neighborhood of zero, and polynomial near the origin with power $c > \frac{n-p+2a-d-1}{2}$.

This result is more substantial than typical convergence rate results for DA algorithms and Gibbs samplers in the sense that it applies to a huge class of mixing densities, whereas typical results apply to relatively small parametric families of Markov chains (see, e.g., Pal and Khare, 2014). Note that, outside of the polynomial case, the only regularity condition on h in Theorem 1 is the rather weak requirement that $\int_0^\infty u^{\frac{d}{2}} h(u) du < \infty$. Thus, for example, Theorem 1 implies that if h is generalized inverse Gaussian, log-normal, inverted gamma (with shape parameter larger than $d/2$), or Fréchet (with shape parameter larger than $d/2$), then the DA Markov chain converges at a geometric rate.

Another notable consequence of Theorem 1 is the following. Suppose that h satisfies the conditions of Theorem 1, and let $B > 0$. Note that we can alter h on the set $[B, \infty)$ in any way we like, and, as long as condition \mathcal{M} continues to hold, the corresponding Markov chain will still be geometrically ergodic.

When h is polynomial near the origin, there is an extra regularity condition for geometric ergodicity that can be somewhat restrictive. For example, take the case where h is the gamma density with shape and rate both equal to $\nu/2$ (so the error density is Student's t with ν degrees of freedom). In this case, Theorem 1 implies that the DA Markov chain will converge at a geometric rate as long as $\nu > n - p + 2a - d + 1$. If $n - p + 2a - d + 1$ is small, then this condition is not too troublesome. However, if this number happens to be large, then Theorem 1 applies only when the degrees of freedom of the t distribution are large, which is not very useful. It is an open question whether the condition $c > \frac{n-p+2a-d-1}{2}$ is necessary.

A couple of special cases of Theorem 1 have appeared previously in the literature. In particular, the result for the gamma mixing density described above was established by Roy and Hobert (2010) in the special case of the independence Jeffreys prior where $a = (d + 1)/2$. Also, Jung and Hobert (2014) showed that, when $d = 1$ and the mixing density is inverted gamma with shape parameter larger than $1/2$, the Markov operator associated with the DA Markov chain is a *trace-class* operator, which implies that the corresponding chain converges at a geometric rate. (Qin and Hobert (2016) provide a substantial generalization of the results in Jung and Hobert (2014).)

It is often possible to improve a given DA algorithm by adding an extra step in between the two conditional draws, and the result is called a *sandwich algorithm* (see, e.g., Khare and Hobert, 2011). The PX-DA and Haar PX-DA algorithms (Liu and Wu, 1999; Meng and van Dyk, 1999; van Dyk and Meng, 2001) are sandwich algorithms with specific recipes for constructing the extra step. Specializing to our context, note that the sampling mechanism of the DA algorithm can be represented as follows: $(\beta, \Sigma) \rightarrow z \rightarrow (\beta', \Sigma')$. The first step is a draw from $\pi(z|\beta, \Sigma, y)$, and the second step is a draw from $\pi(\beta', \Sigma'|z, y)$. Suppose now that $R(z, dz')$ is any Markov transition function that has invariant density

$$\pi(z|y) := \int_{\mathcal{S}_d} \int_{\mathbb{R}^{p \times d}} \pi(\beta, \Sigma, z|y) d\beta d\Sigma .$$

The sandwich algorithm has three steps: $(\beta, \Sigma) \rightarrow z \rightarrow z' \rightarrow (\beta', \Sigma')$. The first and third steps are exactly the same as the DA algorithm, and at the middle step, we draw z' from $R(z, dz')$. Roy and Hobert (2010) constructed the Haar PX-DA version of $R(z, dz')$ for the special case in which $a = \frac{d+1}{2}$. While it is possible to extend their arguments to the case in which $a \neq \frac{d+1}{2}$, an additional regularity condition on h is required. In particular, the Haar PX-DA algorithm can be defined only when

$$\int_0^\infty t^{n + \frac{(d+1-2a)d}{2} - 1} \left[\prod_{i=1}^n h(tz_i) \right] dt < \infty \quad (4)$$

for (almost) all $z \in \mathbb{R}_+^n$. (Condition (4) need only hold for the particular value of (n, d, a) associated with the posterior density at hand.) An argument similar to one used in Roy and Hobert (2010, Section 3) shows that (4) holds if

$$\int_0^\infty u^{\frac{(d+1-2a)d}{2}} h(u) du < \infty . \quad (5)$$

Note that (5) *always* holds when $a = \frac{d+1}{2}$. Now assume that (4) holds, and define a parametric family of density functions, indexed by $z \in \mathbb{R}_+^n$, that take the form

$$\xi(v; z) \propto v^{n + \frac{(d+1-2a)d}{2} - 1} \left[\prod_{i=1}^n h(vz_i) \right] I_{\mathbb{R}_+}(v) .$$

As with the parametric family $\psi(\cdot; s)$, when h is a standard density, ξ often turns out to be standard as well. For example, if h is gamma, inverted gamma, or generalized inverse Gaussian, then ξ turns

out to be a member of the same parametric family. If the current state of the Haar PX-DA Markov chain is $(\beta_m^*, \Sigma_m^*) = (\beta, \Sigma)$, then we simulate the new state, $(\beta_{m+1}^*, \Sigma_{m+1}^*)$, using the following four-step procedure.

Iteration $m + 1$ of the Haar PX-DA algorithm:

1. Draw $\{Z'_i\}_{i=1}^n$ independently with $Z'_i \sim \psi\left(\cdot; (\beta^T x_i - y_i)^T \Sigma^{-1} (\beta^T x_i - y_i)\right)$, and call the result $z' = (z'_1, \dots, z'_n)$.
 2. Draw $V \sim \xi(\cdot; z')$, call the result v , and set $z = (vz'_1, \dots, vz'_n)^T$.
 3. Draw

$$\Sigma_{m+1}^* \sim \text{IW}_d\left(n - p + 2a - d - 1, \left(y^T Q^{-1} y - \mu^T \Omega^{-1} \mu\right)^{-1}\right).$$
 4. Draw $\beta_{m+1}^* \sim \text{N}_{p,d}(\mu, \Omega, \Sigma_{m+1}^*)$
-

Note that the only difference between this algorithm and the DA algorithm is one extra univariate draw (from $\xi(\cdot; \cdot)$) per iteration. Hence, the two algorithms are virtually equivalent from a computational standpoint. Theoretically, the Haar PX-DA algorithm is at least as good as the DA algorithm, both in terms of convergence rate (operator norm) and asymptotic efficiency (Hobert and Marchev, 2008; Khare and Hobert, 2011; Liu and Wu, 1999). The following corollary to Theorem 1 is proven in the Appendix.

Corollary 1. *Assume that (N1) and (N2) hold. Let h be a mixing density that satisfies condition \mathcal{M} as well as (4). If any one of the following conditions holds, then the Haar PX-DA Markov chain is geometrically ergodic.*

1. h is zero near the origin.
2. h is strictly positive in a neighborhood of zero, and faster than polynomial near the origin.
3. h is strictly positive in a neighborhood of zero, and polynomial near the origin with power $c > \frac{n-p+2a-d-1}{2}$.

The remainder of this paper is organized as follows. Section 2 contains a brief description of the latent data model that leads to the DA algorithm, as well as a formal definition of the DA Markov chain. Section 3 contains a drift and minorization analysis of Φ that culminates in a simple sufficient condition for geometric ergodicity that depends only on h . This result is used to prove Theorem 1 in Section 4. In Section 5, we consider the implications of Theorem 1 when h is a member of one of the standard parametric families, and we also develop conditions under which a mixture of mixing

densities leads to a geometric DA Markov chain. In Section 6, we construct and study closed-form upper bounds on the total variation distance to stationarity for the DA Markov chain in the special case where h is an inverted gamma density with shape $(d + 1)/2$ and unit scale. Finally, the Appendix contains the definitions of the inverse Wishart (IW_d) and matrix normal ($\text{N}_{p,d}$) densities, as well as a proof of Corollary 1.

2 The latent data model and the DA Markov chain

In order to formally define the Markov chain that the DA algorithm simulates, we must introduce the latent data model. Suppose that, conditional on (β, Σ) , $\{(Y_i, Z_i)\}_{i=1}^n$ are iid pairs such that

$$Y_i | Z_i = z_i \sim \text{N}_d(\beta^T x_i, \Sigma / z_i)$$

$$Z_i \sim h .$$

Denote the joint density of $\{(Y_i, Z_i)\}_{i=1}^n$ by $\tilde{f}(y, z | \beta, \Sigma)$. It's easy to see that

$$\int_{\mathbb{R}_+^n} \tilde{f}(y, z | \beta, \Sigma) dz = f(y | \beta, \Sigma) ,$$

where the right-hand side is the joint density of the data defined at (3). Now define a (possibly improper) density on $\mathbb{R}^{p \times d} \times \mathcal{S}_d \times \mathbb{R}_+^n$ as follows

$$\pi(\beta, \Sigma, z | y) = \tilde{f}(y, z | \beta, \Sigma) \omega(\beta, \Sigma) ,$$

and note that

$$\int_{\mathbb{R}_+^n} \pi(\beta, \Sigma, z | y) dz = f(y | \beta, \Sigma) \omega(\beta, \Sigma) . \quad (6)$$

It follows that $\pi(\beta, \Sigma, z | y)$ is a proper density if and only if the posterior distribution is proper. Importantly, whether $\pi(\beta, \Sigma, z | y)$ is proper or not, conditions (N1) and (N2) guarantee that the corresponding ‘‘conditional’’ densities, $\pi(\beta, \Sigma | z, y)$ and $\pi(z | \beta, \Sigma, y)$, are well-defined. Indeed, $\pi(\beta, \Sigma | z, y) = \pi(\beta | \Sigma, z, y) \pi(\Sigma | z, y)$, and routine calculations show that $\pi(\beta | \Sigma, z, y)$ is a matrix normal density, and $\pi(\Sigma | z, y)$ is an inverse Wishart density. (The precise forms of these densities can be gleaned from the algorithm stated in the Introduction.) It is also straightforward to show that

$$\pi(z | \beta, \Sigma, y) = \prod_{i=1}^n \psi(z_i; r_i) ,$$

where $r_i = (\beta^T x_i - y_i)^T \Sigma^{-1} (\beta^T x_i - y_i)$ for $i = 1, 2, \dots, n$.

The DA algorithm simulates the Markov chain $\Phi = \{(\beta_m, \Sigma_m)\}_{m=0}^\infty$, whose state space is $\mathcal{X} := \mathbb{R}^{p \times d} \times \mathcal{S}_d$, and whose Markov transition density (Mtd)

$$k(\beta, \Sigma | \tilde{\beta}, \tilde{\Sigma}) = \int_{\mathbb{R}_+^n} \pi(\beta, \Sigma | z, y) \pi(z | \tilde{\beta}, \tilde{\Sigma}, y) dz .$$

We suppress dependence on the data, y , since it is fixed throughout. Note that $\pi(\beta, \Sigma|z, y)$ and $\pi(z|\beta, \Sigma, y)$ are both strictly positive on $Z = \{z \in \mathbb{R}_+ : h(z) > 0\}$, and Z has positive Lebesgue measure. Therefore, $k(\beta, \Sigma|\tilde{\beta}, \tilde{\Sigma})$ is strictly positive on $\mathbb{X} \times \mathbb{X}$, which implies irreducibility and aperiodicity. It's easy to see that (6) is an invariant density for Φ . Consequently, if the posterior is proper, then the chain's invariant density is the target posterior, $\pi^*(\beta, \Sigma|y)$, and the chain is positive recurrent. In fact, it is positive Harris recurrent (because k is strictly positive).

We end this section by describing an interesting simplification that occurs in the special case where $a = (d + 1)/2$ and $n = p + d$. Roy and Hobert (2010) show that when $a = (d + 1)/2$, we have

$$\pi(z|y) = \int_{\mathcal{S}_d} \int_{\mathbb{R}^{p \times d}} \pi(\beta, \Sigma, z|y) d\beta d\Sigma \propto \frac{\prod_{i=1}^n h(z_i)}{|Q|^{\frac{d}{2}} |\Omega|^{\frac{n-p-d}{2}} |\Lambda^T Q^{-1} \Lambda|^{\frac{n-p}{2}}},$$

which is not necessarily integrable in z , because the posterior is not necessarily proper (see, e.g., Fernández and Steel, 1999). (Recall that Q and Ω are both functions of z .) However, when $n = p + d$, Λ is square and non-singular (because of $(N1)$), and we have the stunningly simple formula

$$\pi(z|y) \propto \prod_{i=1}^n h(z_i).$$

Consequently, when $a = (d + 1)/2$ and $n = p + d$, the posterior distribution is proper, and if we are able to draw from the mixing density, h , then we can make an exact draw from the posterior density by drawing sequentially from $\pi(z|y)$, $\pi(\Sigma|z, y)$, and $\pi(\beta|\Sigma, z, y)$, and then ignoring z .

In the next section, we develop a condition on h that implies geometric ergodicity of the DA Markov chain, Φ .

3 A drift and minorization analysis of Φ

Assume that the posterior distribution is proper. Then the DA Markov chain, Φ , is *geometrically ergodic* if there exist $M : \mathbb{X} \rightarrow [0, \infty)$ and $\rho \in [0, 1)$ such that, for all $m \in \mathbb{N}$,

$$\int_{\mathcal{S}_d} \int_{\mathbb{R}^{p \times d}} \left| k^m(\beta, \Sigma|\tilde{\beta}, \tilde{\Sigma}) - \pi^*(\beta, \Sigma|y) \right| d\beta d\Sigma \leq M(\tilde{\beta}, \tilde{\Sigma}) \rho^m, \quad (7)$$

where k^m is the m -step Mtd. The quantity on the left-hand side of (7) is, of course, the total variation distance between the posterior distribution and the distribution of (β_m, Σ_m) conditional on $(\beta_0, \Sigma_0) = (\tilde{\beta}, \tilde{\Sigma})$. We will establish that (7) holds using the so-called drift/minorization method. (For background, see Jones and Hobert (2001) and Roberts and Rosenthal (2004).) We begin by describing exactly what drift and minorization are in the context of our problem.

Drift condition: Let $V : \mathbb{R}^{p \times d} \times \mathcal{S}_d \rightarrow [0, \infty)$, and assume that there exist $\lambda \in [0, 1)$ and $L \in \mathbb{R}$ such that

$$\int_{\mathcal{S}_d} \int_{\mathbb{R}^{p \times d}} V(\beta, \Sigma) k(\beta, \Sigma|\tilde{\beta}, \tilde{\Sigma}) d\beta d\Sigma \leq \lambda V(\tilde{\beta}, \tilde{\Sigma}) + L$$

for all $(\tilde{\beta}, \tilde{\Sigma}) \in \mathbb{R}^{p \times d} \times \mathcal{S}_d$.

Minorization condition: Fix $l > 0$ and define $B_l = \{(\beta, \Sigma) : V(\beta, \Sigma) \leq l\}$. Assume that there exist $\epsilon \in [0, 1)$ and a probability density function $f^* : \mathbb{R}^{p \times d} \times \mathcal{S}_d \rightarrow [0, \infty)$ (both of which depend on l), such that

$$k(\beta, \Sigma | \tilde{\beta}, \tilde{\Sigma}) \geq \epsilon f^*(\beta, \Sigma)$$

for all $(\tilde{\beta}, \tilde{\Sigma}) \in B_l$.

The following result, which follows directly from Theorem 12 in Rosenthal (1995), provides a bound on the total variation distance in terms of the drift and minorization conditions.

Theorem 2 (Rosenthal, 1995). *Suppose that Φ satisfies a drift condition and a minorization condition for some $l > 2L/(1 - \lambda)$. Fix the starting value $\Phi_0 = (\tilde{\beta}, \tilde{\Sigma})$, and define two constants as follows*

$$\alpha = \frac{1 + l}{1 + 2L + \lambda l} \quad \text{and} \quad U = 1 + 2(\lambda l + L).$$

Then for any $0 < r < 1$,

$$\int_{\mathcal{S}_d} \int_{\mathbb{R}^{p \times d}} \left| k^m(\beta, \Sigma | \tilde{\beta}, \tilde{\Sigma}) - \pi^*(\beta, \Sigma | y) \right| d\beta d\Sigma \leq (1 - \epsilon)^{rm} + \left(\frac{U^r}{\alpha^{1-r}} \right)^m \left(1 + \frac{L}{1 - \lambda} + V(\tilde{\beta}, \tilde{\Sigma}) \right). \quad (8)$$

Note that the right-hand side of (8) decreases as λ and L decrease, and as ϵ increases. Note also that (8) implies that

$$\begin{aligned} \int_{\mathcal{S}_d} \int_{\mathbb{R}^{p \times d}} \left| k^m(\beta, \Sigma | \tilde{\beta}, \tilde{\Sigma}) - \pi^*(\beta, \Sigma | y) \right| d\beta d\Sigma \\ \leq \left[\max \left\{ (1 - \epsilon)^r, \frac{U^r}{\alpha^{1-r}} \right\} \right]^m \left(2 + \frac{L}{1 - \lambda} + V(\tilde{\beta}, \tilde{\Sigma}) \right), \end{aligned}$$

which has the same form as (7). Thus, Rosenthal's result provides us with a constructive method of establishing geometric ergodicity. Indeed, if one can establish a drift condition, and an associated minorization condition (with $l > 2L/(1 - \lambda)$), then the chain is geometrically ergodic. This is exactly how we will prove Theorem 1.

Here is the main result of this section.

Proposition 1. *Let h be a mixing density that satisfies condition \mathcal{M} . Suppose that there exist $\lambda \in [0, \frac{1}{n-p+2a-1})$ and $L \in \mathbb{R}$ such that*

$$\frac{\int_0^\infty u^{\frac{d-2}{2}} e^{-\frac{su}{2}} h(u) du}{\int_0^\infty u^{\frac{d}{2}} e^{-\frac{su}{2}} h(u) du} \leq \lambda s + L \quad (9)$$

for every $s \geq 0$. Then the posterior distribution is proper, and the DA Markov chain is geometrically ergodic.

Proof. We begin by noting that the drift and minorization technique is applicable whether the posterior distribution is proper or not. (In more technical terms, it is not necessary to demonstrate that the Markov chain under study is positive recurrent before applying the technique.) Moreover, the DA Markov chain cannot be geometrically ergodic if the posterior is improper (since the corresponding chain is not positive recurrent). Hence, conditions that imply geometric ergodicity of the DA Markov chain simultaneously imply propriety of the corresponding posterior distribution.

Our drift function, $V : \mathbb{R}^{p \times d} \times \mathcal{S}_d \rightarrow \mathbb{R}_+$, is as follows

$$V(\beta, \Sigma) = \sum_{i=1}^n (y_i - \beta^T x_i)^T \Sigma^{-1} (y_i - \beta^T x_i) .$$

We choose to work with this particular drift function because it is precisely the drift function that was used by Roy and Hobert (2010) to establish geometric ergodicity of the DA algorithm in the special case where the mixing density is gamma (and $a = \frac{d+1}{2}$). Indeed, since steps 2. and 3. of the DA algorithm do not depend on the mixing density, we are able to recycle some of their drift calculations.

Part I: Minorization. Fix $l > 0$ and define

$$B_l = \{(\beta, \Sigma) : V(\beta, \Sigma) \leq l\} .$$

We will construct $\epsilon \in (0, 1)$ and a density function $f^* : \mathbb{R}^{p \times d} \times \mathcal{S}_d \rightarrow [0, \infty)$ (both of which depend on l) such that, for all $(\tilde{\beta}, \tilde{\Sigma}) \in B_l$,

$$k(\beta, \Sigma | \tilde{\beta}, \tilde{\Sigma}) \geq \epsilon f^*(\beta, \Sigma) .$$

This is the minorization condition. We note that it suffices to construct $\epsilon \in (0, 1)$ and a density function $\hat{f} : \mathbb{R}_+^n \rightarrow [0, \infty)$ such that, for all $(\tilde{\beta}, \tilde{\Sigma}) \in B_l$,

$$\pi(z | \tilde{\beta}, \tilde{\Sigma}, y) \geq \epsilon \hat{f}(z) .$$

Indeed, if such an \hat{f} exists, then for all $(\tilde{\beta}, \tilde{\Sigma}) \in B_l$, we have

$$k(\beta, \Sigma | \tilde{\beta}, \tilde{\Sigma}) = \int_{\mathbb{R}_+^n} \pi(\beta, \Sigma | z, y) \pi(z | \tilde{\beta}, \tilde{\Sigma}, y) dz \geq \epsilon \int_{\mathbb{R}_+^n} \pi(\beta, \Sigma | z, y) \hat{f}(z) dz = \epsilon f^*(\beta, \Sigma) .$$

We now build \hat{f} . Define $\tilde{r}_i = (y_i - \tilde{\beta}^T x_i)^T \tilde{\Sigma}^{-1} (y_i - \tilde{\beta}^T x_i)$, and note that

$$\pi(z | \tilde{\beta}, \tilde{\Sigma}, y) = \prod_{i=1}^n \psi(z_i; \tilde{r}_i) = \prod_{i=1}^n b(\tilde{r}_i) z_i^{\frac{d}{2}} e^{-\frac{\tilde{r}_i z_i}{2}} h(z_i) .$$

Now, for any $s \geq 0$, we have

$$b(s) = \frac{1}{\int_0^\infty u^{\frac{d}{2}} e^{-\frac{su}{2}} h(u) du} \geq \frac{1}{\int_0^\infty u^{\frac{d}{2}} h(u) du} .$$

By definition, if $(\tilde{\beta}, \tilde{\Sigma}) \in B_l$, then $\sum_{i=1}^n \tilde{r}_i \leq l$, which implies that $\tilde{r}_i \leq l$ for each $i = 1, \dots, n$. Thus, if $(\tilde{\beta}, \tilde{\Sigma}) \in B_l$, then for each $i = 1, \dots, n$, we have

$$z_i^{\frac{d}{2}} e^{-\frac{\tilde{r}_i z_i}{2}} h(z_i) \geq z_i^{\frac{d}{2}} e^{-\frac{l z_i}{2}} h(z_i).$$

Therefore,

$$\begin{aligned} \pi(z|\tilde{\beta}, \tilde{\Sigma}, y) &\geq \left[\int_0^\infty u^{\frac{d}{2}} h(u) du \right]^{-n} \prod_{i=1}^n z_i^{\frac{d}{2}} e^{-\frac{l z_i}{2}} h(z_i) \\ &= \left[\frac{\int_0^\infty u^{\frac{d}{2}} e^{-\frac{l u}{2}} h(u) du}{\int_0^\infty u^{\frac{d}{2}} h(u) du} \right]^n \prod_{i=1}^n \frac{z_i^{\frac{d}{2}} e^{-\frac{l z_i}{2}} h(z_i)}{\int_0^\infty u^{\frac{d}{2}} e^{-\frac{l u}{2}} h(u) du} \\ &=: \epsilon \hat{f}(z). \end{aligned} \tag{10}$$

Hence, our minorization condition is established.

Part II: Drift. To establish the required drift condition, we need to bound the expectation of $V(\beta_{m+1}, \Sigma_{m+1})$ given that $(\beta_m, \Sigma_m) = (\tilde{\beta}, \tilde{\Sigma})$. This expectation is given by

$$\begin{aligned} &\int_{\mathcal{S}_d} \int_{\mathbb{R}^{p \times d}} V(\beta, \Sigma) k(\beta, \Sigma | \tilde{\beta}, \tilde{\Sigma}) d\beta d\Sigma \\ &= \int_{\mathbb{R}_+^n} \left\{ \int_{\mathcal{S}_d} \left[\int_{\mathbb{R}^{p \times d}} V(\beta, \Sigma) \pi(\beta | \Sigma, z, y) d\beta \right] \pi(\Sigma | z, y) d\Sigma \right\} \pi(z | \tilde{\beta}, \tilde{\Sigma}, y) dz. \end{aligned}$$

Calculations in Roy and Hobert's (2010) Section 4 show that

$$\int_{\mathcal{S}_d} \left[\int_{\mathbb{R}^{p \times d}} V(\beta, \Sigma) \pi(\beta | \Sigma, z, y) d\beta \right] \pi(\Sigma | z, y) d\Sigma \leq (n - p + 2a - 1) \sum_{i=1}^n \frac{1}{z_i}.$$

It follows from (9) that

$$\begin{aligned} &\int_{\mathbb{R}_+^n} \left\{ \int_{\mathcal{S}_d} \left[\int_{\mathbb{R}^{p \times d}} V(\beta, \Sigma) \pi(\beta | \Sigma, z, y) d\beta \right] \pi(\Sigma | z, y) d\Sigma \right\} \pi(z | \tilde{\beta}, \tilde{\Sigma}, y) dz \\ &\leq (n - p + 2a - 1) \int_{\mathbb{R}_+^n} \left[\sum_{i=1}^n \frac{1}{z_i} \right] \pi(z | \tilde{\beta}, \tilde{\Sigma}, y) dz \\ &= (n - p + 2a - 1) \sum_{i=1}^n b(\tilde{r}_i) \int_0^\infty u^{\frac{d-2}{2}} e^{-\frac{\tilde{r}_i u}{2}} h(u) du \\ &\leq (n - p + 2a - 1) \left(\lambda \sum_{i=1}^n \tilde{r}_i + nL \right) \\ &= \lambda(n - p + 2a - 1) V(\tilde{\beta}, \tilde{\Sigma}) + (n - p + 2a - 1)nL \\ &= \lambda' V(\tilde{\beta}, \tilde{\Sigma}) + L', \end{aligned}$$

where $\lambda' := \lambda(n - p + 2a - 1) \in [0, 1)$ and $L' := (n - p + 2a - 1)nL$. Since the minorization condition holds for any $l > 0$, an appeal to Theorem 2 yields the result. This completes the proof. \square

Remark 1. A straightforward argument shows that, if the mixing density $h(u)$ satisfies the conditions of Proposition 1, then so does every member of the corresponding scale family given by $\frac{1}{\sigma}h(\frac{u}{\sigma})$, for $\sigma > 0$.

In the next section, we parlay Proposition 1 into a proof of Theorem 1. The key is to show that h satisfies (9) as long as it converges to zero at the origin at an appropriate rate.

4 Proof of Theorem 1

In this section, we prove three corollaries, which, taken together, constitute Theorem 1. There is one corollary for each of the three classes of mixing densities defined in the Introduction. Before we begin stating and proving these corollaries, we define some notation that will be used throughout the remainder of the paper. Fix $\lambda \in [0, \infty)$, and let $\mathcal{A}(\lambda)$ denote the set of mixing densities, h , that are strictly positive in a neighborhood of 0, and for which there exists a constant, k_λ , such that

$$\frac{\int_0^\infty \frac{1}{\sqrt{u}} e^{-\frac{su}{2}} h(u) du}{\int_0^\infty \sqrt{u} e^{-\frac{su}{2}} h(u) du} \leq \lambda s + k_\lambda$$

for every $s \geq 0$. For each mixing density, h , that is strictly positive in a neighborhood of 0, define

$$\lambda_h = \inf \{ \lambda \in [0, \infty) : h \in \mathcal{A}(\lambda) \} .$$

If h is not in $\mathcal{A}(\lambda)$ for any $\lambda \in [0, \infty)$, then we set $\lambda_h = \infty$. Here is an example. Suppose that h is a Gamma($\alpha, 1$) density. If $\alpha > 1/2$, then routine calculations show that

$$\frac{\int_0^\infty \frac{1}{\sqrt{u}} e^{-\frac{su}{2}} h(u) du}{\int_0^\infty \sqrt{u} e^{-\frac{su}{2}} h(u) du} = \frac{1}{2\alpha - 1} s + \frac{2}{2\alpha - 1} . \quad (11)$$

So, in this case, $\lambda_h = \frac{1}{2\alpha - 1}$. On the other hand, if $\alpha \in (0, 1/2]$, then $\lambda_h = \infty$.

4.1 Case I: Zero near the origin

Corollary 2. Let h be a mixing density that satisfies condition \mathcal{M} . If h is zero near the origin, then the posterior distribution is proper and the DA Markov chain is geometrically ergodic.

Proof. Fix $s \geq 0$, and recall that $h(u) = 0$ for $u \in (0, \delta)$ for some $\delta > 0$. Hence,

$$\frac{\int_0^\infty u^{\frac{d-2}{2}} e^{-\frac{su}{2}} h(u) du}{\int_0^\infty u^{\frac{d}{2}} e^{-\frac{su}{2}} h(u) du} = \frac{\int_\delta^\infty \frac{1}{\sqrt{u}} u^{\frac{d-1}{2}} e^{-\frac{su}{2}} h(u) du}{\int_\delta^\infty \sqrt{u} u^{\frac{d-1}{2}} e^{-\frac{su}{2}} h(u) du} \leq \frac{\frac{1}{\sqrt{\delta}} \int_\delta^\infty u^{\frac{d-1}{2}} e^{-\frac{su}{2}} h(u) du}{\sqrt{\delta} \int_\delta^\infty u^{\frac{d-1}{2}} e^{-\frac{su}{2}} h(u) du} = \frac{1}{\delta} .$$

Thus, the conditions of Proposition 1 are satisfied and the proof is complete. \square

4.2 Case II: Polynomial near the origin

Our first result in this section shows that λ_h is determined solely by the behavior of the density h near 0.

Lemma 1. *Suppose that h and \tilde{h} are two mixing densities that are both strictly positive in a neighborhood of zero. If*

$$\lim_{u \rightarrow 0} \frac{h(u)}{\tilde{h}(u)} \in (0, \infty),$$

then, $\lambda_h = \lambda_{\tilde{h}}$.

Proof. Assume that $\lambda_{\tilde{h}} < \infty$. We will show that $\lambda_h \leq \lambda_{\tilde{h}}$. Fix $\lambda \in (\lambda_{\tilde{h}}, \infty)$ arbitrarily. Let $\lambda^* = (\lambda_{\tilde{h}} + \lambda)/2$. Since $\lim_{u \rightarrow 0} \frac{h(u)}{\tilde{h}(u)} \in (0, \infty)$, there exists $\eta > 0$ such that

$$C_{1,\eta} < \frac{h(u)}{\tilde{h}(u)} < C_{2,\eta} \quad (12)$$

for every $u \in (0, \eta]$, where $C_{1,\eta}, C_{2,\eta} \in \mathbb{R}_+$ satisfy $\frac{C_{2,\eta}}{C_{1,\eta}} = \sqrt{\frac{\lambda}{\lambda^*}} > 1$. Also, note that for such an η ,

$$\frac{\int_{\eta}^{\infty} \sqrt{u} e^{-\frac{su}{2}} \tilde{h}(u) du}{\int_{\mathbb{R}_+} \sqrt{u} e^{-\frac{su}{2}} \tilde{h}(u) du} \leq \frac{e^{-\frac{s\eta}{2}} \int_{\eta}^{\infty} \sqrt{u} \tilde{h}(u) du}{\int_0^{\eta/2} \sqrt{u} e^{-\frac{su}{2}} \tilde{h}(u) du} \leq \frac{e^{-\frac{s\eta}{4}} \int_{\eta}^{\infty} \sqrt{u} \tilde{h}(u) du}{\int_0^{\eta/2} \sqrt{u} \tilde{h}(u) du}.$$

Consequently,

$$\frac{\int_{\eta}^{\infty} \sqrt{u} e^{-\frac{su}{2}} \tilde{h}(u) du}{\int_{\mathbb{R}_+} \sqrt{u} e^{-\frac{su}{2}} \tilde{h}(u) du} \rightarrow 0 \text{ as } s \rightarrow \infty,$$

so there exists $s_{\eta} > 0$ such that

$$\frac{\int_0^{\eta} \sqrt{u} e^{-\frac{su}{2}} \tilde{h}(u) du}{\int_{\mathbb{R}_+} \sqrt{u} e^{-\frac{su}{2}} \tilde{h}(u) du} = 1 - \frac{\int_{\eta}^{\infty} \sqrt{u} e^{-\frac{su}{2}} \tilde{h}(u) du}{\int_{\mathbb{R}_+} \sqrt{u} e^{-\frac{su}{2}} \tilde{h}(u) du} \geq \sqrt{\frac{\lambda^*}{\lambda}} \quad (13)$$

for every $s \geq s_{\eta}$. It follows from (12) and (13) that for every $s \geq s_{\eta}$,

$$\begin{aligned} \frac{\int_{\mathbb{R}_+} \frac{1}{\sqrt{u}} e^{-\frac{su}{2}} h(u) du}{\int_{\mathbb{R}_+} \sqrt{u} e^{-\frac{su}{2}} h(u) du} &= \frac{\int_0^{\eta} \frac{1}{\sqrt{u}} e^{-\frac{su}{2}} h(u) du}{\int_{\mathbb{R}_+} \sqrt{u} e^{-\frac{su}{2}} h(u) du} + \frac{\int_{\eta}^{\infty} \frac{1}{\sqrt{u}} e^{-\frac{su}{2}} h(u) du}{\int_{\mathbb{R}_+} \sqrt{u} e^{-\frac{su}{2}} h(u) du} \\ &\leq \frac{\int_0^{\eta} \frac{1}{\sqrt{u}} e^{-\frac{su}{2}} h(u) du}{\int_0^{\eta} \sqrt{u} e^{-\frac{su}{2}} h(u) du} + \frac{1}{\eta} \frac{\int_{\eta}^{\infty} \sqrt{u} e^{-\frac{su}{2}} h(u) du}{\int_{\mathbb{R}_+} \sqrt{u} e^{-\frac{su}{2}} h(u) du} \\ &\leq \frac{C_{2,\eta} \int_0^{\eta} \frac{1}{\sqrt{u}} e^{-\frac{su}{2}} \tilde{h}(u) du}{C_{1,\eta} \int_0^{\eta} \sqrt{u} e^{-\frac{su}{2}} \tilde{h}(u) du} + \frac{1}{\eta} \\ &\leq \sqrt{\frac{\lambda}{\lambda^*}} \sqrt{\frac{\lambda}{\lambda^*}} \frac{\int_0^{\eta} \frac{1}{\sqrt{u}} e^{-\frac{su}{2}} \tilde{h}(u) du}{\int_{\mathbb{R}_+} \sqrt{u} e^{-\frac{su}{2}} \tilde{h}(u) du} + \frac{1}{\eta} \\ &\leq \frac{\lambda}{\lambda^*} \frac{\int_{\mathbb{R}_+} \frac{1}{\sqrt{u}} e^{-\frac{su}{2}} \tilde{h}(u) du}{\int_{\mathbb{R}_+} \sqrt{u} e^{-\frac{su}{2}} \tilde{h}(u) du} + \frac{1}{\eta}. \end{aligned}$$

Since $\tilde{h} \in \mathcal{A}(\lambda^*)$, there exists k such that

$$\frac{\int_{\mathbb{R}_+} \frac{1}{\sqrt{u}} e^{-\frac{su}{2}} h(u) du}{\int_{\mathbb{R}_+} \sqrt{u} e^{-\frac{su}{2}} h(u) du} \leq \frac{\lambda}{\lambda^*} (\lambda^* s + k) + \frac{1}{\eta} = \lambda s + \frac{\lambda}{\lambda^*} k + \frac{1}{\eta} \quad (14)$$

for every $s \geq s_\eta$. Our assumptions imply that $\int_{\mathbb{R}_+} \frac{1}{\sqrt{u}} \tilde{h}(u) du < \infty$. Together with (12), this leads to $\int_{\mathbb{R}_+} \frac{1}{\sqrt{u}} h(u) du < \infty$. Then, since

$$\sup_{s \in (0, s_\eta)} \frac{\int_{\mathbb{R}_+} \frac{1}{\sqrt{u}} e^{-\frac{su}{2}} h(u) du}{\int_{\mathbb{R}_+} \sqrt{u} e^{-\frac{su}{2}} h(u) du} \leq \sup_{s \in (0, s_\eta)} \frac{\int_{\mathbb{R}_+} \frac{1}{\sqrt{u}} h(u) du}{e^{-\frac{s}{2}} \int_0^1 \sqrt{u} h(u) du} \leq \frac{e^{\frac{s_\eta}{2}} \int_{\mathbb{R}_+} \frac{1}{\sqrt{u}} h(u) du}{\int_0^1 \sqrt{u} h(u) du},$$

it follows from (14) that $h \in \mathcal{A}(\lambda)$. Hence, $\lambda_h \leq \lambda$. Since $\lambda \in (\lambda_{\tilde{h}}, \infty)$ was arbitrarily chosen, it follows that $\lambda_h \leq \lambda_{\tilde{h}}$.

Now assume that $\lambda_h < \infty$. We can show that $\lambda_{\tilde{h}} \leq \lambda_h$ by noting that

$$\lim_{u \rightarrow 0} \frac{h(u)}{\tilde{h}(u)} \in (0, \infty) \Leftrightarrow \lim_{u \rightarrow 0} \frac{\tilde{h}(u)}{h(u)} \in (0, \infty),$$

and reversing the roles of h and \tilde{h} in the above argument. We have shown that $\lambda_h < \infty$ if and only if $\lambda_{\tilde{h}} < \infty$, and when they are finite, they are equal. \square

Corollary 3. *Let h be a mixing density that satisfies condition \mathcal{M} . If h is polynomial near the origin with power $c > \frac{n-p+2a-d-1}{2}$, then the posterior distribution is proper and the DA Markov chain is geometrically ergodic.*

Proof. We can write

$$\frac{\int_0^\infty u^{\frac{d-2}{2}} e^{-\frac{su}{2}} h(u) du}{\int_0^\infty u^{\frac{d}{2}} e^{-\frac{su}{2}} h(u) du} = \frac{\int_0^\infty \frac{1}{\sqrt{u}} e^{-\frac{su}{2}} h^*(u) du}{\int_0^\infty \sqrt{u} e^{-\frac{su}{2}} h^*(u) du}, \quad (15)$$

where $h^*(u)$ is the mixing density that is proportional to $u^{\frac{d-1}{2}} h(u)$. It's easy to see that h^* is polynomial near the origin with power $c' > \frac{n-p+2a-2}{2}$. (Note that (N2) implies that $c' > 0$, so the integral in the numerator on the right-hand side of (15) is finite.) Let \tilde{h} be the Gamma($c' + 1, 1$) density, which is clearly polynomial near the origin with power c' . Then,

$$\lim_{u \rightarrow 0} \frac{h^*(u)}{\tilde{h}(u)} = \lim_{u \rightarrow 0} \frac{h^*(u)}{u^{c'}} \frac{u^{c'}}{\tilde{h}(u)} \in (0, \infty).$$

Thus, (11) and Lemma 1 imply that $\lambda_{h^*} = \lambda_{\tilde{h}} = 1/(2c' + 1)$, and the result now follows from Proposition 1 since

$$\lambda_{h^*} = \frac{1}{2c' + 1} < \frac{1}{n - p + 2a - 1}.$$

\square

4.3 Case III: Faster than polynomial near the origin

Lemma 2. *Suppose that h and \tilde{h} are two mixing densities that are both strictly positive in a neighborhood of zero. If there exists $\eta > 0$ such that $\frac{h}{\tilde{h}}$ is a strictly increasing function on $(0, \eta]$, then $\lambda_h \leq \lambda_{\tilde{h}}$.*

Proof. First, fix $s > 0$ and define two densities as follows: $h_{s,\eta}(u) = K_{s,\eta} e^{-\frac{su}{2}} h(u) I_{(0,\eta)}(u)$ and $\tilde{h}_{s,\eta}(u) = \tilde{K}_{s,\eta} e^{-\frac{su}{2}} \tilde{h}(u) I_{(0,\eta)}(u)$, where $K_{s,\eta}$ and $\tilde{K}_{s,\eta}$ are normalizing constants. Since $\frac{h}{\tilde{h}}$ is strictly increasing on $(0, \eta]$, it follows that

$$\frac{h_{s,\eta}(u)}{\tilde{h}_{s,\eta}(u)} > 1 \Leftrightarrow \frac{h(u)}{\tilde{h}(u)} > \frac{\tilde{K}_{s,\eta}}{K_{s,\eta}} \Leftrightarrow u > u^*$$

for some $u^* \in (0, \eta)$. This shows that the densities $\tilde{h}_{s,\eta}$ and $h_{s,\eta}$ cross exactly once in the interval $(0, \eta)$, which is their common support. Let F and \tilde{F} denote the distribution functions associated with $h_{s,\eta}$ and $\tilde{h}_{s,\eta}$, respectively. Note that, for $t \in (0, u^*)$, $\tilde{F}(t) - F(t) = \int_0^t (\tilde{h}_{s,\eta}(u) - h_{s,\eta}(u)) du > 0$, and for $t \in [u^*, \eta)$,

$$\tilde{F}(t) - F(t) = \int_0^t (\tilde{h}_{s,\eta}(u) - h_{s,\eta}(u)) du = - \int_t^\eta (\tilde{h}_{s,\eta}(u) - h_{s,\eta}(u)) du > 0.$$

Hence, $\tilde{F}(t) > F(t)$ for all $t \in (0, \eta)$. It follows that a random variable with density $\tilde{h}_{s,\eta}$ is stochastically dominated by a random variable with density $h_{s,\eta}$. This stochastic dominance implies that

$$\int_0^\eta \frac{1}{\sqrt{u}} \tilde{h}_{s,\eta}(u) du \geq \int_0^\eta \frac{1}{\sqrt{u}} h_{s,\eta}(u) du \quad \text{and} \quad \int_0^\eta \sqrt{u} \tilde{h}_{s,\eta}(u) du \leq \int_0^\eta \sqrt{u} h_{s,\eta}(u) du. \quad (16)$$

Now define two more densities as follows

$$h_\eta(u) = \frac{h(u)}{\int_0^\eta h(v) dv} I_{(0,\eta)}(u) \quad \text{and} \quad \tilde{h}_\eta(u) = \frac{\tilde{h}(u)}{\int_0^\eta \tilde{h}(v) dv} I_{(0,\eta)}(u).$$

It follows from (16) that

$$\frac{\int_{\mathbb{R}_+} \frac{1}{\sqrt{u}} e^{-\frac{su}{2}} \tilde{h}_\eta(u) du}{\int_{\mathbb{R}_+} \sqrt{u} e^{-\frac{su}{2}} \tilde{h}_\eta(u) du} = \frac{\int_{\mathbb{R}_+} \frac{1}{\sqrt{u}} \tilde{h}_{s,\eta}(u) du}{\int_{\mathbb{R}_+} \sqrt{u} \tilde{h}_{s,\eta}(u) du} \geq \frac{\int_{\mathbb{R}_+} \frac{1}{\sqrt{u}} h_{s,\eta}(u) du}{\int_{\mathbb{R}_+} \sqrt{u} h_{s,\eta}(u) du} = \frac{\int_{\mathbb{R}_+} \frac{1}{\sqrt{u}} e^{-\frac{su}{2}} h_\eta(u) du}{\int_{\mathbb{R}_+} \sqrt{u} e^{-\frac{su}{2}} h_\eta(u) du}.$$

Hence, $\lambda_{h_\eta} \leq \lambda_{\tilde{h}_\eta}$. Since

$$\lim_{u \rightarrow 0} \frac{h(u)}{h_\eta(u)} = \int_0^\eta h(v) dv \in \mathbb{R}_+ \quad \text{and} \quad \lim_{u \rightarrow 0} \frac{\tilde{h}(u)}{\tilde{h}_\eta(u)} = \int_0^\eta \tilde{h}(v) dv \in \mathbb{R}_+,$$

it follows from Lemma 1 that $\lambda_h = \lambda_{h_\eta}$ and $\lambda_{\tilde{h}} = \lambda_{\tilde{h}_\eta}$. □

Corollary 4. *Let h be a mixing density that satisfies condition \mathcal{M} . If h is faster than polynomial near the origin, then the posterior distribution is proper and the DA Markov chain is geometrically ergodic.*

Proof. Again, define $h^*(u)$ to be the mixing density that is proportional to $u^{\frac{d-1}{2}}h(u)$. In light of (15), it suffices to show that $\lambda_{h^*} = 0$. First, note that h^* is faster than polynomial near the origin. Fix $c > 0$ and define $\tilde{h}(u) = (c+1)u^c I_{(0,1)}(u)$. Clearly, $\lambda_{\tilde{h}} = \frac{1}{2c+1}$. Since h^* is faster than polynomial near the origin, there exists $\eta_c \in (0, 1)$ such that $\frac{h^*(u)}{\tilde{h}(u)}$ is strictly increasing in $(0, \eta_c)$. Thus, Lemma 2 implies that $\lambda_{h^*} \leq \lambda_{\tilde{h}} = \frac{1}{2c+1}$. But c was arbitrary, so $\lambda_{h^*} = 0$. The result now follows immediately from Proposition 1. \square

Taken together, Corollaries 2, 3 and 4 are equivalent to Theorem 1. Hence, our proof of Theorem 1 is complete.

5 Examples and a result concerning mixtures of mixing densities

We claimed in the Introduction that every mixing density which is a member of a standard parametric family is either polynomial near the origin, or faster than polynomial near the origin. Here we provide some details. When we write $W \sim \text{Gamma}(\alpha, \gamma)$, we mean that W has density proportional to $w^{\alpha-1}e^{-w\gamma}I_{\mathbb{R}_+}(w)$. By $W \sim \text{Beta}(\alpha, \gamma)$, we mean that the density is proportional to $w^{\alpha-1}(1-w)^{\gamma-1}I_{(0,1)}(w)$, and by $W \sim \text{Weibull}(\alpha, \gamma)$, we mean that the density is proportional to $w^{\alpha-1}e^{-\gamma w^\alpha}I_{\mathbb{R}_+}(w)$. In all three cases, we need $\alpha, \gamma > 0$. It is clear that these densities are all polynomial near the origin with $c = \alpha - 1$. Moreover, condition \mathcal{M} always holds. Hence, according to Theorem 1, if the mixing density is $\text{Gamma}(\alpha, \gamma)$, $\text{Beta}(\alpha, \gamma)$ or $\text{Weibull}(\alpha, \gamma)$ with $\alpha > \frac{n-p+2a-d+1}{2}$, then the DA Markov chain is geometrically ergodic.

By $W \sim F(\nu_1, \nu_2)$, we mean that W has density proportional to

$$\frac{w^{(\nu_1-2)/2}}{\left(1 + \left(\frac{\nu_1}{\nu_2}\right)w\right)^{(\nu_1+\nu_2)/2}} I_{\mathbb{R}_+}(w),$$

where $\nu_1, \nu_2 > 0$. These densities are polynomial near the origin with $c = (\nu_1 - 2)/2$. To get a geometrically ergodic chain in this case, we need $\nu_1 > n - p + 2a - d + 1$ and $\nu_2 > d$. (The second condition is to ensure that condition \mathcal{M} holds.) Consider the shifted Pareto family with density given by

$$\frac{\gamma\alpha^\gamma}{(w+\alpha)^{\gamma+1}} I_{\mathbb{R}_+}(w),$$

where $\alpha, \gamma > 0$. This density is polynomial near the origin with $c = 0$. Since the requirement that $c > \frac{n-p+2a-d-1}{2}$ forces c to be strictly positive, Theorem 1 is not applicable to this family.

By $W \sim \text{IG}(\alpha, \gamma)$, we mean that W has density proportional to $w^{-\alpha-1} e^{-\gamma/w} I_{\mathbb{R}_+}(w)$, where $\alpha, \gamma > 0$. For any $c > 0$, the derivative of $\log(h(w)/w^c)$ is

$$\frac{-(\alpha + c + 1)}{w} + \frac{\gamma}{w^2} = \frac{1}{w} \left[-(\alpha + c + 1) + \frac{\gamma}{w} \right],$$

which is clearly strictly positive in a neighborhood of zero. Hence, the $\text{IG}(\alpha, \gamma)$ densities are all faster than polynomial near the origin. Thus, Theorem 1 implies that, as long as $\alpha > d/2$, the DA Markov chain is geometrically ergodic.

By $W \sim \text{GIG}(v, a, b)$, we mean that W has a generalized inverse Gaussian distribution with density given by

$$h(w) = \frac{1}{2K_v(\sqrt{ab})} \left(\frac{a}{b}\right)^{\frac{v}{2}} w^{v-1} \exp\left\{-\frac{1}{2}\left(aw + \frac{b}{w}\right)\right\} I_{\mathbb{R}_+}(w), \quad (17)$$

where $a, b \in \mathbb{R}_+$ and $v \in \mathbb{R}$. Taking $v = -\frac{1}{2}$ leads to the standard inverse Gaussian density (with a nonstandard parametrization). By $W \sim \text{Log-normal}(\mu, \gamma)$, we mean that W has density proportional to

$$\frac{1}{w} \exp\left\{-\frac{1}{2\gamma}(\log w - \mu)^2\right\} I_{\mathbb{R}_+}(w),$$

where $\mu \in \mathbb{R}$ and $\gamma > 0$. By $W \sim \text{Fréchet}(\alpha, \gamma)$, we mean that W has density proportional to

$$w^{-(\alpha+1)} e^{-\frac{\gamma}{w^\alpha}} I_{\mathbb{R}_+}(w),$$

where $\alpha, \gamma > 0$. Arguments similar to those used in the inverted gamma case above show that all members of these three families are faster than polynomial near the origin. Moreover, condition \mathcal{M} holds for all the Log-normal and GIG densities, and for all Fréchet(α, γ) densities with $\alpha > d/2$. Thus, the corresponding DA Markov chains are all geometric.

We end this section with a result concerning mixtures of mixing densities.

Proposition 2. *Let I be an index set equipped with a probability measure ξ . Consider a family of mixing densities $\{h_a\}_{a \in I}$ such that $\lambda_{h_a} = 0$ for every $a \in I$. In particular, for every $a \in I$ and every $\lambda \in (0, 1)$, there exists $k_{a,\lambda} > 0$ such that*

$$\frac{\int_0^\infty \frac{1}{\sqrt{u}} e^{-\frac{su}{2}} h_a(u) du}{\int_0^\infty \sqrt{u} e^{-\frac{su}{2}} h_a(u) du} \leq \lambda s + k_{a,\lambda}$$

for every $s \geq 0$. Suppose that, for every $\lambda \in (0, 1)$,

$$\sup_{a \in I} k_{a,\lambda} < \infty. \quad (18)$$

Then $\lambda_h = 0$ where $h(u) = \int_I h_a(u) \xi(da)$.

Proof. Fix $\lambda \in (0, 1)$. For every $s \geq 0$, we have

$$\begin{aligned}
\frac{\int_0^\infty \frac{1}{\sqrt{u}} e^{-\frac{su}{2}} h(u) du}{\int_0^\infty \sqrt{u} e^{-\frac{su}{2}} h(u) du} &= \frac{\int_I \left(\int_0^\infty \frac{1}{\sqrt{u}} e^{-\frac{su}{2}} h_a(u) du \right) \xi(da)}{\int_0^\infty \sqrt{u} e^{-\frac{su}{2}} h(u) du} \\
&\leq \frac{\int_I (\lambda s + k_{a,\lambda}) \left(\int_0^\infty \sqrt{u} e^{-\frac{su}{2}} h_a(u) du \right) \xi(da)}{\int_0^\infty \sqrt{u} e^{-\frac{su}{2}} h(u) du} \\
&\leq \left(\lambda s + \sup_{a \in I} k_{a,\lambda} \right) \frac{\int_I \int_0^\infty \sqrt{u} e^{-\frac{su}{2}} h_a(u) du \xi(da)}{\int_0^\infty \sqrt{u} e^{-\frac{su}{2}} h(u) du} \\
&= \lambda s + \sup_{a \in I} k_{a,\lambda} .
\end{aligned}$$

Since this holds for all $\lambda \in (0, 1)$, the result follows. \square

Remark 2. *If the index set, I , in Proposition 2 is a finite set, then (18) is automatically satisfied.*

Here's a simple application of Proposition 2.

Proposition 3. *Let $\{h_i\}_{i=1}^M$ be a finite set of mixing densities that all satisfy condition \mathcal{M} , and are all either zero near the origin, or faster than polynomial near the origin. Define*

$$h(u) = \sum_{i=1}^M w_i h_i(u) ,$$

where $w_i > 0$ and $\sum_{i=1}^M w_i = 1$. Then the posterior distribution is proper and the DA Markov chain is geometrically ergodic.

Proof. Since Proposition 2 implies that $\lambda_h = 0$, the arguments in the proof of Corollary 4 can be applied to prove the result. \square

6 Exact total variation bounds when h is inverted gamma

Our main results (Theorem 1 & Corollary 1) are *qualitative* in the sense that they provide conditions under which certain Markov chains are geometrically ergodic, but they *do not* provide computable bounds on the total variation distance to stationarity. In this section, we demonstrate that, for fixed h , with a bit more work, we can sometimes get such computable bounds via Theorem 2. We assume throughout this section that the mixing density is $\text{IG}\left(\frac{d+1}{2}, 1\right)$ because the required calculations are particularly simple in this case. Indeed, there is a closed form upper bound on the total variation distance to stationarity, and we can study how this upper bound behaves as a function of n , p , d , and the data.

Note that the right-hand side of (8) depends on λ and L from the drift condition, ϵ and l from the minorization condition, as well as the drift function evaluated at the starting value. We now develop closed form expressions for ϵ , λ , and L .

Recall from (10) that

$$\epsilon^{\frac{1}{n}} = \frac{\int_0^\infty u^{\frac{d}{2}} e^{-\frac{lu}{2}} h(u) du}{\int_0^\infty u^{\frac{d}{2}} h(u) du}.$$

A simple calculation shows that the denominator is $\sqrt{\pi}/\Gamma(\frac{d+1}{2})$, and (17) shows that the numerator equals

$$\frac{2K_{-\frac{1}{2}}(\sqrt{2}l)}{\Gamma(\frac{d+1}{2})} \left(\frac{l}{2}\right)^{\frac{1}{4}}.$$

Now since $K_{-\frac{1}{2}}(u) = K_{\frac{1}{2}}(u) = \sqrt{\frac{\pi}{2u}}e^{-u}$, it follows that $\epsilon = e^{-n\sqrt{2}l}$. Note that ϵ does not depend on p, d , nor on the actual data that is observed. On the negative side, when l is fixed, ϵ goes to zero exponentially fast as n gets large. While the user does have some control over l , it cannot be chosen any smaller than $2L/(1 - \lambda)$.

We now develop an explicit drift function. First, using (17) and the fact that $K_{-\frac{3}{2}}(u) = K_{\frac{3}{2}}(u) = \sqrt{\frac{\pi}{2u}}e^{-u}(1 + \frac{1}{u})$, we have

$$\frac{\int_0^\infty u^{\frac{d-2}{2}} e^{-\frac{su}{2}} h(u) du}{\int_0^\infty u^{\frac{d}{2}} e^{-\frac{su}{2}} h(u) du} = \frac{\sqrt{s}}{\sqrt{2}} + \frac{1}{2}.$$

Now fix $B > 0$, and note that

$$\frac{\sqrt{s}}{\sqrt{2}} + \frac{1}{2} \leq \frac{1}{2B}s + \frac{B}{4} + \frac{1}{2},$$

for all $s \geq 0$. Plugging this into the drift argument at the end of Section 3, we have

$$\begin{aligned} & \int_{S_d} \int_{\mathbb{R}^{p \times d}} V(\beta, \Sigma) k(\beta, \Sigma | \tilde{\beta}, \tilde{\Sigma}) d\beta d\Sigma \\ & \leq (n - p + 2a - 1) \sum_{i=1}^n \left[\frac{\int_0^\infty u^{\frac{d-2}{2}} e^{-\frac{\tilde{r}_i u}{2}} h(u) du}{\int_0^\infty u^{\frac{d}{2}} e^{-\frac{\tilde{r}_i u}{2}} h(u) du} \right] \\ & = (n - p + 2a - 1) \sum_{i=1}^n \left[\frac{\sqrt{\tilde{r}_i}}{\sqrt{2}} + \frac{1}{2} \right] \\ & \leq (n - p + 2a - 1) \sum_{i=1}^n \left[\frac{1}{2B} \tilde{r}_i + \frac{B}{4} + \frac{1}{2} \right] \\ & = \frac{(n - p + 2a - 1)}{2B} \sum_{i=1}^n \tilde{r}_i + \frac{n(n - p + 2a - 1)}{4} (B + 2) \\ & = \frac{(n - p + 2a - 1)}{2B} V(\tilde{\beta}, \tilde{\Sigma}) + \frac{n(n - p + 2a - 1)}{4} (B + 2) \\ & = \lambda V(\tilde{\beta}, \tilde{\Sigma}) + L. \end{aligned}$$

Here, B is chosen so that $\lambda < 1$. Suppose that B is independent of (p, d) . Then λ and L are both free of d , and neither depends explicitly on the data. (Of course, if we use the default prior,

$a = (d + 1)/2$, then both will depend on d , and increasing d will make things worse.) Both also decrease with p . Note that in order to have $\lambda < 1$, as $n \rightarrow \infty$, B must be at least $O(n)$. Suppose that $B = O(n)$, then $\lambda = O(1)$, while L increases with n cubically. If B goes to infinity faster than n , then λ can go to 0, but L goes to infinity faster than $O(n^3)$.

Acknowledgments. The authors thank two anonymous referees for helpful comments and suggestions. The first author was supported by NSF Grants DMS-11-06395 & DMS-15-11945, and the third by NSF Grants DMS-11-06084 & DMS-15-11945.

University of Florida
 Department of Statistics
 Gainesville, FL 32611
 USA
 jhobert@stat.ufl.edu

Appendices

A Matrix Normal and Inverse Wishart Densities

Matrix Normal Distribution Suppose Z is an $r \times c$ random matrix with density

$$f_Z(z) = \frac{1}{(2\pi)^{\frac{rc}{2}} |A|^{\frac{c}{2}} |B|^{\frac{r}{2}}} \exp \left[-\frac{1}{2} \text{tr} \left\{ A^{-1} (z - \theta) B^{-1} (z - \theta)^T \right\} \right],$$

where θ is an $r \times c$ matrix, A and B are $r \times r$ and $c \times c$ positive definite matrices. Then Z is said to have a *matrix normal distribution* and we denote this by $Z \sim \mathbf{N}_{r,c}(\theta, A, B)$ (Arnold, 1981, Chapter 17).

Inverse Wishart Distribution Suppose W is an $r \times r$ random positive definite matrix with density

$$f_W(w) = \frac{|w|^{-\frac{m+r+1}{2}} \exp \left\{ -\frac{1}{2} \text{tr}(\Theta^{-1} w^{-1}) \right\}}{2^{\frac{mr}{2}} \pi^{\frac{r(r-1)}{4}} |\Theta|^{\frac{m}{2}} \prod_{i=1}^r \Gamma\left(\frac{1}{2}(m+1-i)\right)} I_{\mathcal{S}_r}(W),$$

where $m > r - 1$ and Θ is an $r \times r$ positive definite matrix. Then W is said to have an *inverse Wishart distribution* and this is denoted by $W \sim \text{IW}_r(m, \Theta)$.

B Proof of Corollary 1

Let $\Phi = \{(\beta_m, \Sigma_m)\}_{m=0}^{\infty}$ and $\Phi^* = \{(\beta_m^*, \Sigma_m^*)\}_{m=0}^{\infty}$ denote the DA and Haar PX-DA Markov chains, respectively. In both cases, the state space is $\mathbf{X} := \mathbb{R}^{p \times d} \times \mathcal{S}_d$.

Define a Hilbert space consisting of complex valued functions on X that are square integrable and have mean zero with respect to $\pi^*(\beta, \Sigma|y)$, namely

$$L_0^2(\pi) := \left\{ f \in L^2(\pi) \mid \int_{\mathcal{S}_d} \int_{\mathbb{R}^{p \times d}} f(\beta, \Sigma) \pi^*(\beta, \Sigma|y) d\beta d\Sigma = 0 \right\},$$

where

$$L^2(\pi) := \left\{ f : \mathsf{X} \rightarrow \mathbb{C} \mid \int_{\mathcal{S}_d} \int_{\mathbb{R}^{p \times d}} |f(\beta, \Sigma)|^2 \pi^*(\beta, \Sigma|y) d\beta d\Sigma < \infty \right\}.$$

For $f, g \in L_0^2(\pi)$, their inner product is given by

$$\langle f, g \rangle_\pi = \int_{\mathcal{S}_d} \int_{\mathbb{R}^{p \times d}} f(\beta, \Sigma) \overline{g(\beta, \Sigma)} \pi^*(\beta, \Sigma|y) d\beta d\Sigma,$$

and $\|f\| = \sqrt{\langle f, f \rangle_\pi}$.

The DA and Haar PX-DA Markov chains are each associated with an operator on $L_0^2(\pi)$. Indeed, let $k(\beta, \Sigma|\tilde{\beta}, \tilde{\Sigma})$ and $k^*(\beta^*, \Sigma^*|\tilde{\beta}^*, \tilde{\Sigma}^*)$ denote the Mtds of the DA and Haar PX-DA chains, respectively. Let $K : L_0^2(\pi) \rightarrow L_0^2(\pi)$ be the operator that maps $f \in L_0^2(\pi)$ into

$$(Kf)(\tilde{\beta}, \tilde{\Sigma}) = \int_{\mathcal{S}_d} \int_{\mathbb{R}^{p \times d}} f(\beta, \Sigma) k(\beta, \Sigma|\tilde{\beta}, \tilde{\Sigma}) d\beta d\Sigma,$$

and define $K^* : L^2(\pi) \rightarrow L^2(\pi)$ analogously using k^* in place of k . Now let $L_{0,1}^2(\pi)$ denote the subset of $L_0^2(\pi)$ consisting of functions with norm 1. The norm of the operator K is defined as

$$\|K\| = \sup_{f \in L_{0,1}^2(\pi)} \|Kf\|,$$

and the norm of K^* is defined analogously.

The DA and Haar PX-DA Markov chains are both reversible (Hobert and Marchev, 2008). It follows that, for these chains, geometric ergodicity is equivalent to having an operator norm strictly less than 1 (Kontoyiannis and Meyn, 2012; Roberts and Rosenthal, 1997). Since the hypotheses of Corollary 1 contain those of Theorem 1, it follows from Theorem 1 that $\|K\| < 1$. Finally, by combining Hobert and Marchev's (2008) Corollary 1, Proposition 1 and Proposition 4, we arrive at the conclusion that $\|K^*\| \leq \|K\|$. Hence, $\|K^*\| \leq \|K\| < 1$, so the Haar PX-DA Markov chain is geometrically ergodic.

References

- ALA-LUHTALA, J. and PICHÉ, R. (2016). Gaussian scale mixture models for robust linear multivariate regression with missing data. *Communications in Statistics - Simulation and Computation*, **45** 791–813.

- ANDREWS, D. F. and MALLOWS, C. L. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society, Series B*, **36** 99–102.
- ARNOLD, S. F. (1981). *The Theory of Linear Models and Multivariate Analysis*. Wiley, New York.
- CHOY, S. T. B., CHAN, J. S. K. and MAKOV, U. E. (2016). Robust Bayesian analysis of loss reserving data using scale mixtures distributions. *Journal of Applied Statistics*, **43** 396–411.
- FERNÁNDEZ, C. and STEEL, M. F. J. (1999). Multivariate Student-t regression models: Pitfalls and inference. *Biometrika*, **86** 153–167.
- FERNÁNDEZ, C. and STEEL, M. F. J. (2000). Bayesian regression analysis with scale mixtures of normals. *Econometric Theory*, **16** 80–101.
- FERREIRA, M. A. and SALAZAR, E. (2014). Bayesian reference analysis for exponential power regression models. *Journal of Statistical Distributions and Applications*, **1**.
- FLEGAL, J. M., HARAN, M. and JONES, G. L. (2008). Markov chain Monte Carlo: Can we trust the third significant figure? *Statistical Science*, **23** 250–260.
- HOBERT, J. P. and MARCHEV, D. (2008). A theoretical comparison of the data augmentation, marginal augmentation and PX-DA algorithms. *The Annals of Statistics*, **36** 532–554.
- JONES, G. L. and HOBERT, J. P. (2001). Honest exploration of intractable probability distributions via Markov chain Monte Carlo. *Statistical Science*, **16** 312–34.
- JUNG, Y. J. and HOBERT, J. P. (2014). Spectral properties of MCMC algorithms for Bayesian linear regression with generalized hyperbolic errors. *Statistics & Probability Letters*, **95** 92–100.
- KHARE, K. and HOBERT, J. P. (2011). A spectral analytic comparison of trace-class data augmentation algorithms and their sandwich variants. *The Annals of Statistics*, **39** 2585–2606.
- KONTOYIANNIS, I. and MEYN, S. P. (2012). Geometric ergodicity and the spectral gap of non-reversible Markov chains. *Probability Theory and Related Fields*, **154** 327–339.
- ŁATUSZYŃSKI, K., MIASOJEDOW, B. and NIEMIRO, W. (2013). Nonasymptotic bounds on the estimation error of MCMC algorithms. *Bernoulli*, **19** (5A) 2033–2066.
- LEE, N., CHOI, H. and KIM, S.-H. (2016). Bayes shrinkage estimation for high-dimensional VAR models with scale mixture of normal distributions for noise. *Computational Statistics & Data Analysis*, **101** 250–276.
- LIU, C. (1996). Bayesian robust multivariate linear regression with incomplete data. *Journal of the American Statistical Association*, **91** 1219–1227.

- LIU, J. S. and WU, Y. N. (1999). Parameter expansion for data augmentation. *Journal of the American Statistical Association*, **94** 1264–1274.
- MENG, X.-L. and VAN DYK, D. A. (1999). Seeking efficient data augmentation schemes via conditional and marginal augmentation. *Biometrika*, **86** 301–320.
- NEVO, D. and RITOV, Y. (2016). On Bayesian robust regression with diverging number of predictors. *Electronic Journal of Statistics*, **10** 3045–3062.
- PAL, S. and KHARE, K. (2014). Geometric ergodicity for Bayesian shrinkage models. *Electronic Journal of Statistics*, **8** 604–645.
- QIN, Q. and HOBERT, J. P. (2016). Trace-class Monte Carlo Markov chains for Bayesian multivariate linear regression with non-Gaussian errors. Tech. rep., University of Florida. ArXiv:1602.00136v1.
- ROBERTS, G. O. and ROSENTHAL, J. S. (1997). Geometric ergodicity and hybrid Markov chains. *Electronic Communications in Probability*, **2** 13–25.
- ROBERTS, G. O. and ROSENTHAL, J. S. (1998). Markov chain Monte Carlo: Some practical implications of theoretical results (with discussion). *Canadian Journal of Statistics*, **26** 5–31.
- ROBERTS, G. O. and ROSENTHAL, J. S. (2004). General state space Markov chains and MCMC algorithms. *Probability Surveys*, **1** 20–71.
- ROSENTHAL, J. S. (1995). Minorization conditions and convergence rates for Markov chain Monte Carlo. *Journal of the American Statistical Association*, **90** 558–566.
- ROY, V. and HOBERT, J. P. (2010). On Monte Carlo methods for Bayesian multivariate regression models with heavy-tailed errors. *Journal of Multivariate Analysis*, **101** 1190–1202.
- SALAZAR, E., FERREIRA, M. A. and MIGON, H. S. (2012). Objective Bayesian analysis for exponential power regression models. *Sankhyā: The Indian Journal of Statistics*, **74** 107–125.
- VAN DYK, D. A. and MENG, X.-L. (2001). The art of data augmentation (with discussion). *Journal of Computational and Graphical Statistics*, **10** 1–50.
- WEST, M. (1984). Outlier models and prior distributions in Bayesian linear regression. *Journal of the Royal Statistical Society, Series B*, **46** 431–439.
- YANG, F. and YUAN, H. (2017). A non-iterative Bayesian sampling algorithm for linear regression models with scale mixtures of normal distributions. *Computational Economics*, **49** 579–597.